DOCUMENT RESUME

ED 373 087 TM 021 976

AUTHOR Hsu, Yaowen; Ackerman, Terry A.
TITLE Equating Reading Test Scores That Combine Narrative and Expository Test Formats.
PUB DATE Apr 94
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994). For related documents, see TM 021 975-977.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Context Effect; Elementary School Students; *Equated Scores; Grade 6; Intermediate Grades; *Reading Tests; Scaling; Test Format; *True Scores
IDENTIFIERS Expository Text; Illinois; *Illinois Goal Assessment Program; Linking Metrics; Narrative Text; Partial Credit Model

ABSTRACT

This paper summarizes an investigation of the format used for equating the 1993 Illinois Goal Assessment Program (IGAP) sixth grade reading test. In 1992, each student took only one test, either a narrative test or an expository test. In 1993, there was only one test, which included both formats. Several possible approaches for linking the 1993 test to the 1992 tests, including use of the partial credit model and true-score equating, are proposed and investigated in this study. The sample size for the 1992 narrative test was 10,178. The expository test sample was 10,277, and the sample for the 1993 test was 4,830. Results show that the 1993 examinees have a higher mean-scaled score than the 1992 examinees if the test is linked to the narrative test, but a lower score if linked to the expository test. Three tables and 10 figures present analysis results. (Contains 8 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 373 087

Equating Reading Test Scores That Combine Narrative and Expository Test Formats

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TERRY ACKERMAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Yaowen Hsu

Terry A. Ackerman

University of Illinois, Urbana-Champaign

Running Head: EQUATING IGAP READING TESTS

Paper presented at the 1994 Annual Meeting of

the American Educational Research Association, New Orleans

TM021976

Abstract

The purpose of this paper is to dummarize an examination of the format that was used for equating the 1993 Illinois Goal Assessment Program's sixth grade reading test for. In 1992 each student took only one test: either a narrative test or an expository test. In 1993 there was only one test that included both types of tests. Several possible approaches for linking the 1993 test to the 1992 tests are proposed and investigated in this study. Results show that the 1993 examinees have higher mean scaled score than 1992 if the test is linked to the narrative test; but lower if linked to the expository test.

Equating Reading Test Scores That Combine Narrative and Expository Test Formats

The Illinois Goal Assessment Program (IGAP) is designed to provide statewide assessment of established state goals created in an educational reform package by the State of Illinois in 1985. The assessment areas include reading, mathematics, writing, science, and social sciences. One of IGAP purposes is to measure student performance relative to state goals, and describe how schools and districts perform compared to a set of established standards. To evaluate progress of schools and districts over time, it is important that tests be properly linked or equated from year to year.

In contrast with older tests that used isolated paragraphs and fragmented text, the passages used in the IGAP reading test are intact pieces of literature, stories, and essays that match classroom reading assignments and typical student reading experiences. Each year, nearly a half million students take IGAP reading tests. The 1993 IGAP reading tests were administered to all public school students enrolled in grades 3, 6, 8, and 10. Each test consisted of two 15-item reading passages. Each test was administered in two 40-minute sessions with a 10-minute break between them. One passage followed an expository (informational) format, the other a narrative (story) format. Each passage had been administered individually in 1992. That is, in 1992 students took only a 15-item reading passage (either narrative or

expository) but in 1993 students took a reading test that contained both narrative and expository passages. The items used in 1993 were idential to those used in 1992.

Thus, from an equating perspective, several possibilities exist for linking the 1993 test results to the 1992 reported scale. Two logical ways include: linking via the 15-item expository passage or linking via the 15-item narrative passage. This paper will suggest and compare several equating procedures using the IGAP reading data focusing only on the results for the grade 6 students.

Background

Each item in the 1993 IGAP Reading test consisted of a question followed by five statements (choices, alternatives). The examinee had to judge if each statement is correct or incorrect on the basis of the text that preceded the items. Students were instructed that there could be one, two, or three true statements for each item. One point was awarded each time the examinee identified the correctness of the statement. Each of 30 items was graded on a zero to five scale. Thus, for the entire test the raw score ranged from zero to 150. These raw scores were then transformed to the reporting scale, which had a mean of 250 and standard deviation of 100 in the first year of testing. Since 1988, the equating of test results has been managed by procedures that

have their roots in classical test theory. However, in 1993, attempts were made to equate using item response theory (IRT). For the 0-5 scoring scheme, the most appropriate IRT model is the partial credit model, particularly because there is a one-to-one correspondence between the number correct score and the θ-scale.

Partial Credit Model

The partial credit model (Master, 1982) is written as

$$\pi_{xi}(\theta_n) = \frac{\exp \sum_{j=0}^{x} (\theta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} (\theta_n - \delta_{ij})} \qquad x = 0, 1, \ldots, m_i$$

where $\pi_{xi}(\theta_n)$ is the probability of examinee n scoring x on the m_i-step item i,

$m_i = 5$, the number of score categories

θ_n is the ability of examinee n,

δ_{ij} is the difficulty of step j in item i,

x is a count of the successfully completed item steps, i.e., the score.

Each step in the partial credit model corresponds to one of six score categories (zero to five). That is, the probability of examinee n getting score x on item i is the function of the examinee's ability and 6 difficulty parameters of score categories of item i. As in the Rasch model, the test score is a

sufficient statistic of the ability parameter.

The computer program BIGSTEPS (Linacre & Wright, 1993) was used to calibrate items, link items, and estimate examinees' ability parameters.

True-score Equating

The true-score equating (Lord, 1980) is used to equate 1992 and 1993 tests. Given test X and test Y, for each test, there is one-to-one (often monotonic increasing, at least for the partial credit model) relationship between the true score (ξ) and the ability (θ), i.e., $\xi = f(\theta)$, called the test characteristic function (TCC). For example, the TCC of the partial credit model can be written as

$$\xi(\theta) = \sum_{i} \sum_{x} \pi_{ix}(\theta)\, x\,.$$

For any specified θ value, there will be a pair of ξ_X and ξ_Y. These pairs provide a one-to-one monotone equating function of X and Y. Figure 1 shows an equating plot with two imaginary TCCs. A true score of 7 in Test Y is equivalent to a true score about 23 in Test X.

Insert Figure 1 about here

The values in Table 20.1 of BIGSTEPS give a TCC used for true-score equating.

Method

In this study, the sample size (the number of examinees) of the 1992 narrative test is 10178, 10277 for the 1992 expository test, and 4830 for the 1993 test.

There are several ways for linking the 93 students' performance to 92 students' performance scale. Different linking can be used because of the special test item structure of 1993 tests. The 1993 tests consist of two passages/subtests, and each passage was a test in 1992. Thus, from an equating perspective, two possibilities existed for linking the 1993 test results to the reporting scale, one via the 15-item expository passage and one via the 15-item narrative passage.

However, there are some problems about comparing test performance of 92 and 93 students. For example, consider examinees A to G, each with a different score patterns as followings:

Examinee	Test	Response vector
A	1992 narrative test	555555555555555
B	1992 expository test	555555555555555
C	1993 entire test	555555555555555555555555555555
D	1993 entire test	555555555555555555555555555550
E	1993 entire test	055555555555555555555555555555
F	1993 entire test	555555555555555xxxxxxxxxxxxxxx

G 1993 entire test xxxxxxxxxxxxxxx555555555555555

where x denotes any possible score from 0 to 5. One question that arises is: who is most able? who is next? and, who is least able? Without a doubt, examinee C in 1993 is one of the most able examinees among these 7 examinees, because all items are answered correctly. However, is examinee A or B who took the 1992 tests as able as examinee C? Is examinee D of 1993 better than A? Is examinee F less able than examine A, if all x's (items that were not taken) are 0s? Finally, who is better, A or B in 1992 (because there is no anchored examinee)? Note that in 93 tests none of students get all 30 items correctly. One cannot determine if A has taken the other 15 items, what proportion of items he or she would have answered correctly. On the other hand, when we say that examinee F is less able than A, F might complain that the additional items made the test more difficult.

It should be reasonable to say that F or G could be the same able as A or B, if the expository (narrative) items that F or G answered incorrectly are easier than narrative (expository) items. And, F could be more able than A if the expository items are very more difficult than narrative ones.

After considering the above ideas, several possible procedures are proposed to equate the 93 test to the 92 tests. These are illustrated in Figure 2 and described below.

Insert Figure 2 about here

Case 1 (anchored on 1992 narrative items):

I. Calibrate the 1992 narrative test.

II. Fixed 1993 narrative items as 1992 narrative item parameter estimates, calibrate 1993 expository item parameters and estimate examinees' abilities ($\hat{\theta}_N$).

III. Equate 1993 raw score to 1992 raw score through $\hat{\theta}_N$, and find scaled scores (SS_N) of 1993 examinees.

Case 2 (anchored 92 expository items):

I. Calibrate 92 expository items using the response data of those examinees took the 1992 expository test.

II. Fixed 93 expository items using 92 expository item parameter estimates, calibrate 93 narrative item parameters and estimate examinees' abilities ($\hat{\theta}_E$).

III. Equate 93 raw scores to 92 raw scores through $\hat{\theta}_E$, and find scaled scores (SS_E) of 1993 examinees.

Case 3 (weighted combination):

I. Obtain $\hat{\theta}_N$ and $\hat{\theta}_E$ by Cases 1 and 2.

II. (1) Find equated scaled scores (SSs) for $\hat{\theta}_N$ and $\hat{\theta}_E$, respectively, and then determine final SSs which is the weighted mean of SS_N and SS_E.

Note: (a) if 92 narrative and 92 expository item parameter estimates are on the same scale, then equal weights can be taken for combination.

(b) Also, weights can be considered as functions of information of ability parameter.

(c) If both estimates are not sure on the same scale, then it is hard to combine θ_N and θ_E.

Results

Table 1 shows means, standard deviations, minima, and maxima of ability estimates of four cases for calibrating/equating 1992 and 1993 narrative/expository tests. Those examinees who took the 1992 narrative test have the mean θ-ability estimate 1.32 and the mean scaled score 246.16. When anchored on this 1992 narrative test, the mean θ-ability estimates and scaled scores of examinees taking the 1993 test are 1.49 and 252.66, respectively, which is .17 higher on the θ-ability scale and 6.5 higher on the scaled score. Those who took the 1992 expository test have mean ability 1.32

and mean scaled score 244.52. The mean θ-ability and scaled score of those examinees taking the 1993 test, when fixed on the 1992 expository items, are 1.26 and 227.20 which are .06 low on the θ-ability scale and 17.32 low on the scaled score. Using equating of Case 3, the mean ability and scaled score of 1993 examinees are 1.37 and 239.93. That is, 1993 examinees is higher on the ability scale but lower on the scaled score than 1992 examinees either taking narrative or expository test. Note, that because examinees taking the narrative test are different from those taking the expository test in 1992, the mean of both tests can not be computed. However, if we consider the standard deviations of abilities (greater than .68) and scaled scores (greater than 83.50), then there may not be significant changes on ability from 1992 to 1993.

Insert Table 1 about here

Figures 3 to 8 show that the 1992 and 1993 examinees have very similar distributions on the ability and the scaled score, suggesting that the the two groups are probably equivalent. For case 1, the ability distribution of examinees taking the 1992 narrative test are nearly normal distributed, as is the ability distribution of examinee score for the 1993 test after linking it to the 1992 narrative test (Figures 3a and 3b). The scaled score distributions of

both tests are also similar and skewed negatively (Figures 4a and 4b). Similar observations for Case 2 (Figures 5a, 5b, 6a and 6b). The ability and scaled score distributions of the 1993 examinees for Case 3 are similar to other cases (Figures 7 and 8). In sum, because of the similarity of the shapes between 1992 equating and 1993 equated tests, the locations of the two θ-ability distributions can be compared.

Insert Figures 3a to 8 about here

Figures 9 and 10 are the plots of true-score equating for Case 1 and Case 2. In each plot, there are two test characteristic curves (TCCs): one is for the 1992 equating test (where test scores are from 0 to 75) and the other is for the 1993 equated test (where test scores are between 0 and 150). In each plot, both TCCs are very close, especially for θ-ability greater than -1. For θ-ability > -.5, two TCCs are almost coincident when the 1992 expository test is used as linking items. In fact, the smallest estimated θ-ability for each test was greater than -1.0. Such a similiality might indicate that the examinee groups of 1992 ad 1993 do not differ significantly, because all items of both 92 tests are the same as of the 93 test.

Insert Figures 9 & 10 about here

Because the two 15-item passages in the 1993 test are the same as those in the 1992 narrative test and the 1992 expository test, the 1993 test and its subtests as well as two 1992 tests can be used to explore the equating procedures proposed by this study.

When calibrating the 1993 test, the mean θ-ability of the 1993 examinee group is 1.22. When only calibrating the 15 items of the 1993 narrative subtest, the mean θ-ability is 1.34. When only calibrating the 15 items of the 1993 expository subtest, the mean ability is 1.27. If, analog to Case 1, the 15 narrative subtest items are calibrated first, next, given these 15 item estimates, the 1993 30 items are calibrated, then the mean ability is 1.29. If, analog to Case 2, the 15 items of the 1993 expository subtest are calibrated first, and then 30 items are calibrated given these expository item estimates, the mean ability is 1.25. This observation implies that if calibration/equating involves narrative items, then the mean ability of 1993 group becomes higher. However, the standard deviations of the above five calibrations are between .63 and .76, so the differences appear to be negligible (see Table 2).

Insert Table 2 about here

Another explanation why the narrative test gives higher ability estimates is because the 1993 examinees perform a little bit better than 1992 examinee group who took the narrative test, but similar to the 1992 expository group. To show this, we equate 1992 tests to the 1993 test (see Table 3). This should give better equating between the narrative and expository tests, because in 1993 all examinees took both tests. Hence, the 1993 test is calibrated first (the mean ability of 1993 examinees is 1.22), then item estimates are used to estimate abilities of two 1992 examinee groups. In this way, the mean ability of 1992 narrative group is 1.07 which is lower than the 1993 group. The mean ability of 1992 expository group is 1.27 that is .05 higher than the 1993 group and .2 higher than the 1992 narrative group.

Insert Table 3 about here

If we concentrated only on 15 narrative items, from the view of 1993 (i.e., first calibrate 1993 15 narrative items and obtain the 1993 examinees' ability estimates, then fixing the item parameters estimate 1992 narrative abilities), the mean ability of 1993 group is .2 higher. From the view of 1992

(i.e., fixing the '93 item parameters to the '92 estimates), the mean ability of 1993 group is .22 higher. If only concentrating on 15 expository items, from the view of 1993 the mean ability of the 1993 group is .04 lower than the 1992 expository group. From the view of 1992, the mean ability of the 1993 group is .02 lower. Again, because standard deviations are around .7, these differences are not significantly.

Discussion

In this study, four approaches are proposed to equate the 1993 IGAP reading test for the sixth grade to their 1992 counterparts. The mean equated scaled score of 1993 is 6.5 higher by Case 1, 17.32 lower by Case 2, 5 lower by Case 3 than 1992. However, such differences are not statistically significant.

Using the partial credit IRT model, results show that the 1993 group performed very similiar to the 1992 expository group but slightly different from the 1992 narrative group. Therefore, each of the above equated results would appear reasonable. For instance, the purpose of Case 1 was to compare the 1993 group with the 1992 narrative group. Case 2's goal was to essentially compare the 1993 group with the 1992 expository group. On the other hand, perhaps observed differences result from multidimensionality of these two subtests, i.e., narrative vs. expository (cf. Bolt & Ackerman, 1994).

However, such differences are not conclusive. If such differences are too large from a school adminstrator's perspective, then other better equating approaches could be considered, such as equating from a multidimensional perspective. Or, if the two passages measure the same ability, then, instead of using equal weights, we can take the weighted mean of scaled scores obtained from Case 1 and Case 2, using test information as weights (See Evans & Ackerman, 1994, for the issue of test/item information).

On the other hand, because all items of the 1993 test were identical to those of the 1992 tests, we may reverse the procedure to cross validate our results. That is, we may first calibrate the 1993 test, then using these item parameter estimates to estimate abilities of the 1992 narrative group and the 1992 expository group. Subsequently one can use the lookup table of ability vs. scaled score made from the 1992 test data to determine the scaled scores of the 1993 group.

References

Bolt, D., & Ackerman, T. (1994, April). An Examination of the influence of expository and narrative passages on the dimensionality of the IGAP reading test. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Evans, J., & Ackerman, T. (1994, April). XXXX. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

The Illinois goal assessment program 1991 technical manual. Springfield: Illinois State Board of Education, 1992.

The Illinois goal assessment program: Information bulletin. Springfield: Illinois State Board of Education, 1992.

Lord, F. (1980). Applications of item response theory to practical testing problems. Hillsdale, New Jersey: LEA.

Masters, G. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

Masters, G. (1988). The analysis of partial credit scoring. Applied Measurement in Education, 1, 279-297.

Linacre, J., & Wright, B. (1993). BIGSTEPS user's guide. MESA Press, Chicago, IL: University of Illinois.

Table 1 Equating Design Results

Case		Year	Mean	Std Dev	Minimum	Maximum
One	Theta	92	1.32	.78	- .83	5.12
		93	1.49	.70	- .89	4.09
	Scaled Score	92	246.16	105.30	8	456
		93	252.66	83.70	8	460
Two	Theta	92	1.32	.72	- .65	4.42
		93	1.26	.67	- .93	3.73
	Scaled Score	92	244.52	105.60	8	461
		93	227.20	86.87	8	452
Three	Theta	92	——	——	——	——
		93	1.37	.68	- .91	3.91
	Scaled Score	92	——	——	——	——
		93	239.93	83.50	8	456

Table 2 Sample Means and Standard Deviations of the Ability Distributionthe of the 1993 Examinee Group By various Calibrations

Estimation of abilities of 1993 examinees	Mean	Std Dev	Mini.	Maxi.
On the 1993 test of 30 items	1.22	.63	-.88	3.60
Theta-NE[1]	1.25	.64	-.88	3.65
Case 2[2]	1.26	.67	-.93	3.73
Only on the 15 items of the 1993 expository subtest	1.27	.70	-.96	4.28
Theta-EN[3]	1.29	.66	-.92	3.75
Using the 1992 expository item estimates	1.30	.75	-1.00	4.43
Only on the 15 items of the 1993 narrative subtest	1.34	.76	-1.10	4.63
Case 3[2]	1.37	.68	-.91	3.91
Case 4[2]	1.39	.72	-.99	4.06
Case 1[2]	1.49	.70	-.89	4.09
Using the 1992 narrative item estimates	1.54	.84	-1.21	5.12

Note: 1. First, obtain the item parameter estimates of the 1993 narrative subtest, then given these item estimates, calibrate the 1993 test.
2. See descriptions in the section Methods.
3. First, obtain the item parameter estimates of the 1993 expository subtest, then given these item estimates, calibrate the 1993 test.

Table 3 Sample Means and Standard Deviations of Ability Distributions for 1992 and 1993 Groups By Various Calibrations

	Mean	Std Dev	Mini.	Maxi.
for the 1993 30-item test	1.22	.63	-.88	3.60
for the 1992 narrative test using 15 narrative item estimates of the 1993 30-item test	1.07	.66	-.68	4.43
for the 1992 expository test using 15 expository item estimates of the 1993 30-item test	1.27	.65	-.43	4.20
for the 1993 15-item narrative subtest	1.34	.76	-1.10	4.63
for the 1992 narrative test using item estimates of the 1993 narrative subtest	1.14	.71	-.75	4.63
for the 1993 15-item expository subtest	1.27	.70	-.96	4.28
for the 1992 expository test using item estimates of the 1993 15-item expository subtest	1.31	.66	-.44	4.28

Figure Caption

Figure 1. An example of true-score equating.

Figure 2. Four equating schemes shown graphically.

Figure 3a. Ability distribution of the 1992 narrative group.

Figure 3b. Ability distribution of the 1993 examinee group by Case 1.

Figure 4a. Scaled score distribution of the 1992 narrative group.

Figure 4b. Scaled score distribution of the 1993 examinee group by Case 1.

Figure 5a. Ability distribution of the 1992 expository group.

Figure 5b. Ability distribution of the 1993 examinee group by Case 2.

Figure 6a. Scaled score distribution of the 1992 expository group.

Figure 6b. Scaled score distribution of the 1993 examinee group by Case 2.

Figure 7. Ability distribution of the 1993 examinee group by Case 3.

Figure 8. Scaled score distribution of the 1993 examinee group by Case 3.

Figure 9. True-score equating plot for Case 1.

Figure 10. True-score equating plot for Case 2.

Figure 1

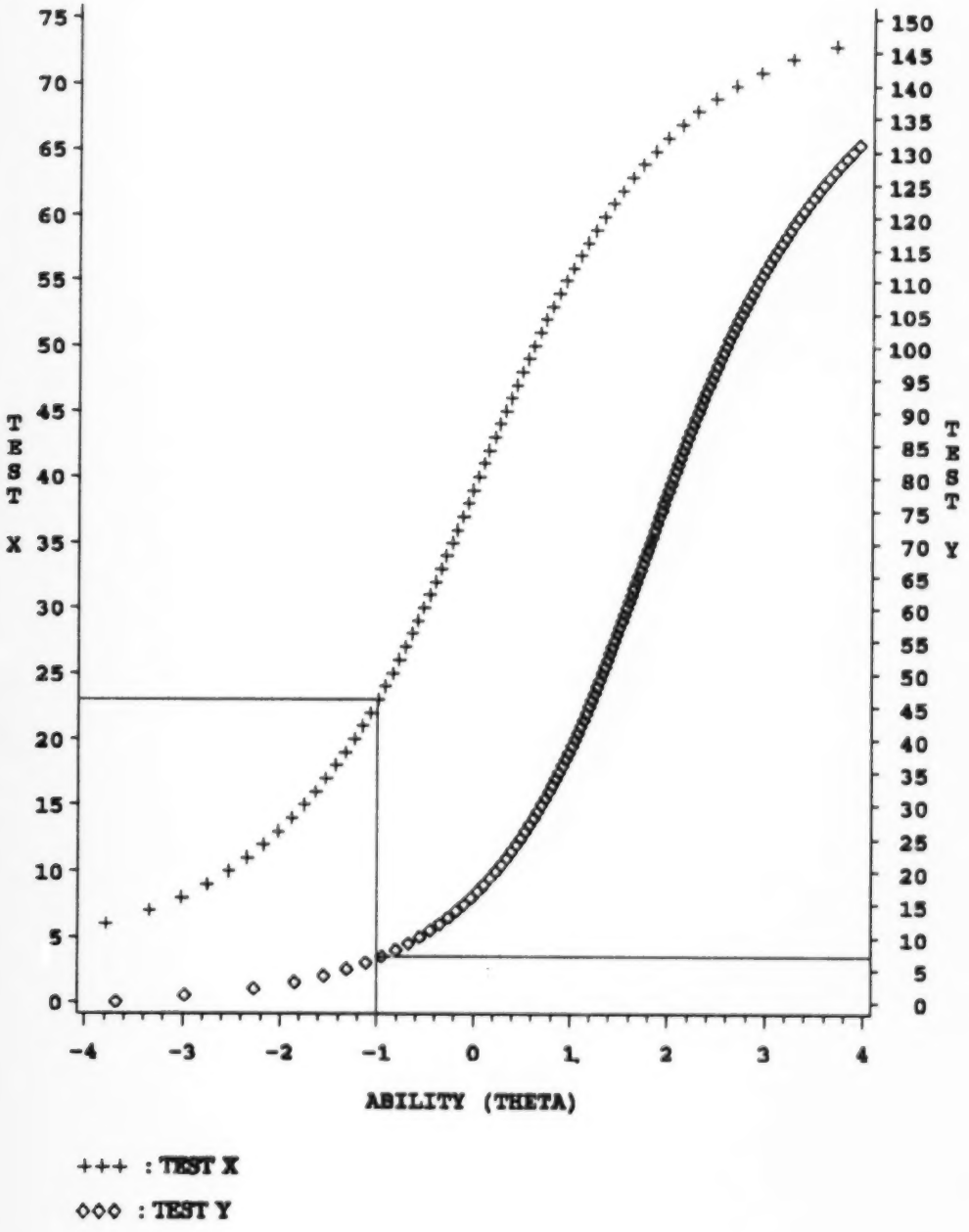
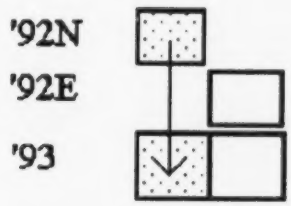
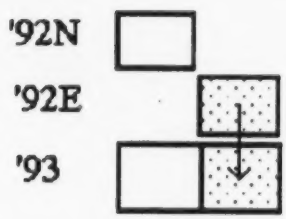


Figure 2

Case 1



Case 2



Case 3

Average of Case 1 and Case 2

Case 4

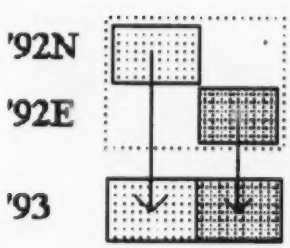


Figure 3a

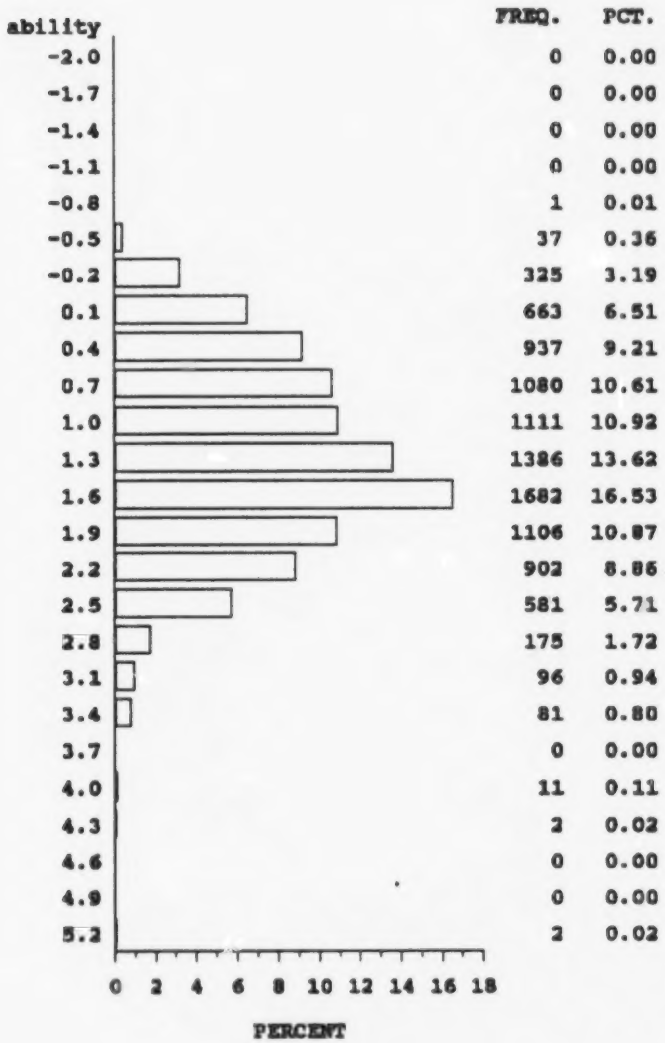

ability
FREQ.
PCT.
-2.0 0 0.00
-1.7 0 0.00
-1.4 0 0.00
-1.1 0 0.00
-0.8 1 0.01
-0.5 37 0.36
-0.2 325 3.19
0.1 663 6.51
0.4 937 9.21
0.7 1080 10.61
1.0 1111 10.92
1.3 1386 13.62
1.6 1682 16.53
1.9 1106 10.87
2.2 902 8.86
2.5 581 5.71
2.8 175 1.72
3.1 96 0.94
3.4 81 0.80
3.7 0 0.00
4.0 11 0.11
4.3 2 0.02
4.6 0 0.00
4.9 0 0.00
5.2 2 0.02
0 2 4 6 8 10 12 14 16 18
PERCENT

Figure 3b

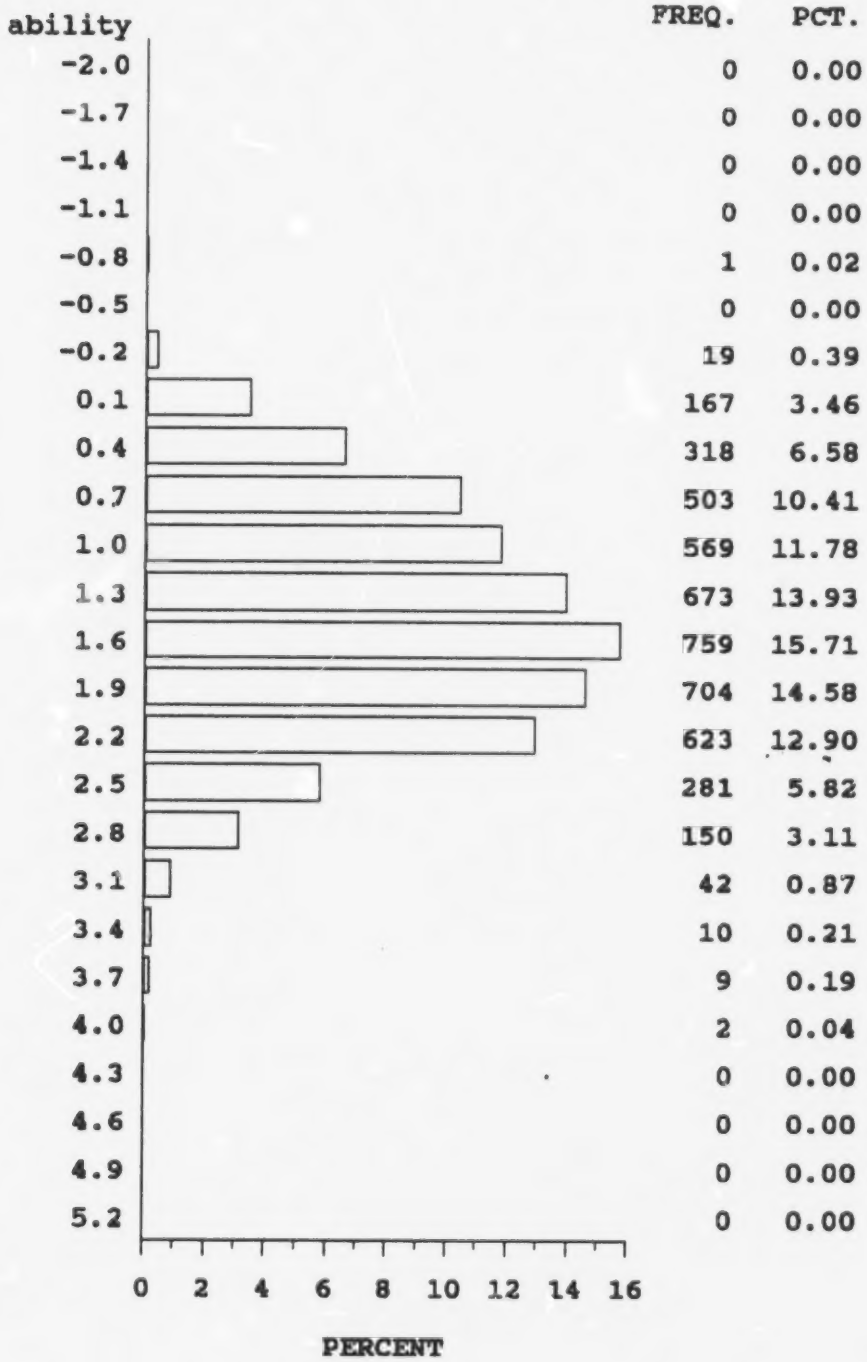

ability
FREQ.
PCT.
-2.0 0 0.00
-1.7 0 0.00
-1.4 0 0.00
-1.1 0 0.00
-0.8 1 0.02
-0.5 0 0.00
-0.2 19 0.39
0.1 167 3.46
0.4 318 6.58
0.7 503 10.41
1.0 569 11.78
1.3 673 13.93
1.6 759 15.71
1.9 704 14.58
2.2 623 12.90
2.5 281 5.82
2.8 150 3.11
3.1 42 0.87
3.4 10 0.21
3.7 9 0.19
4.0 2 0.04
4.3 0 0.00
4.6 0 0.00
4.9 0 0.00
5.2 0 0.00
0 2 4 6 8 10 12 14 16
PERCENT

Figure 4a

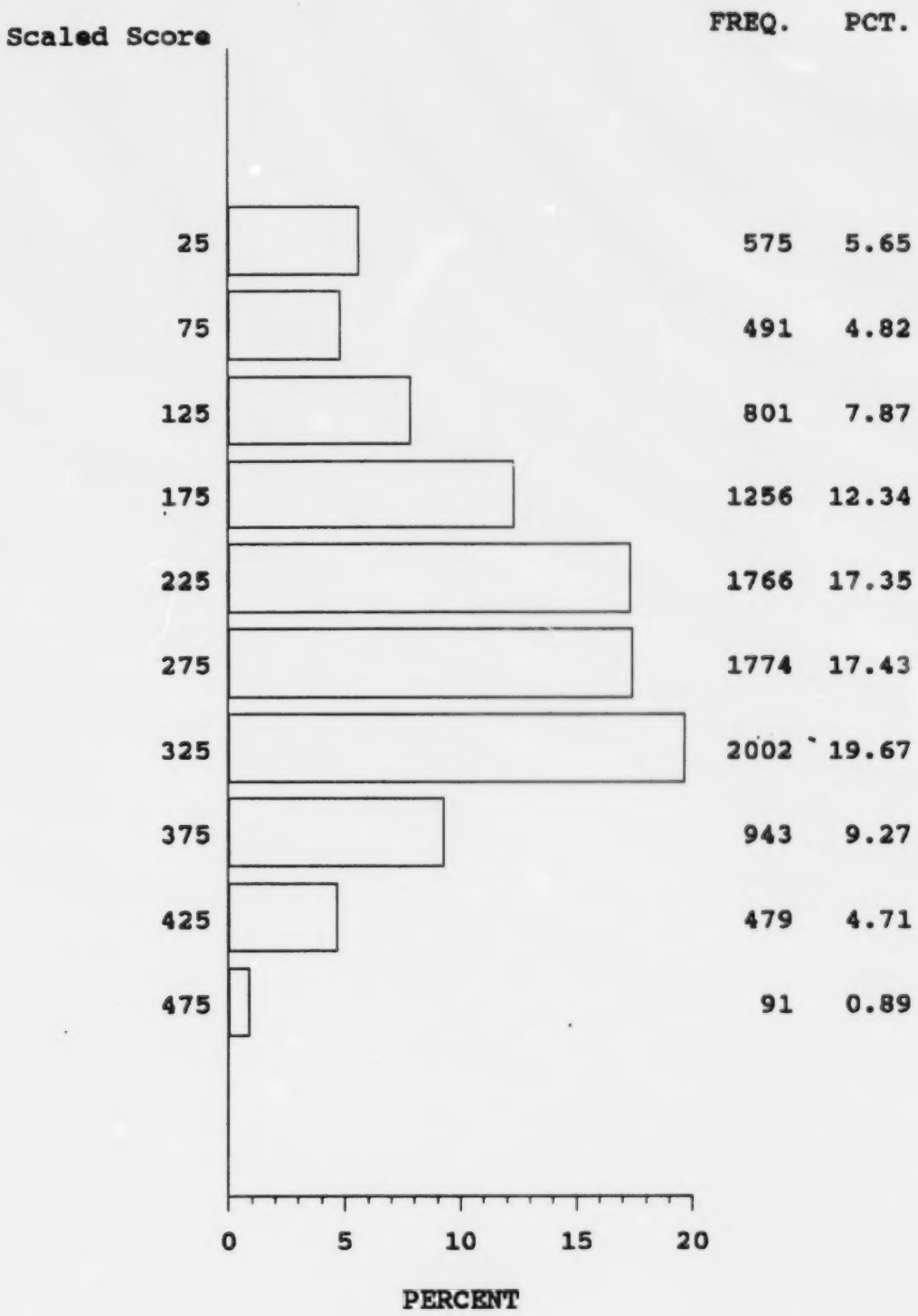

Scaled Score
FREQ.
PCT.
25
575
5.65
75
491
4.82
125
801
7.87
175
1256
12.34
225
1766
17.35
275
1774
17.43
325
2002
19.67
375
943
9.27
425
479
4.71
475
91
0.89
0
5
10
15
20
PERCENT

Figure 4b

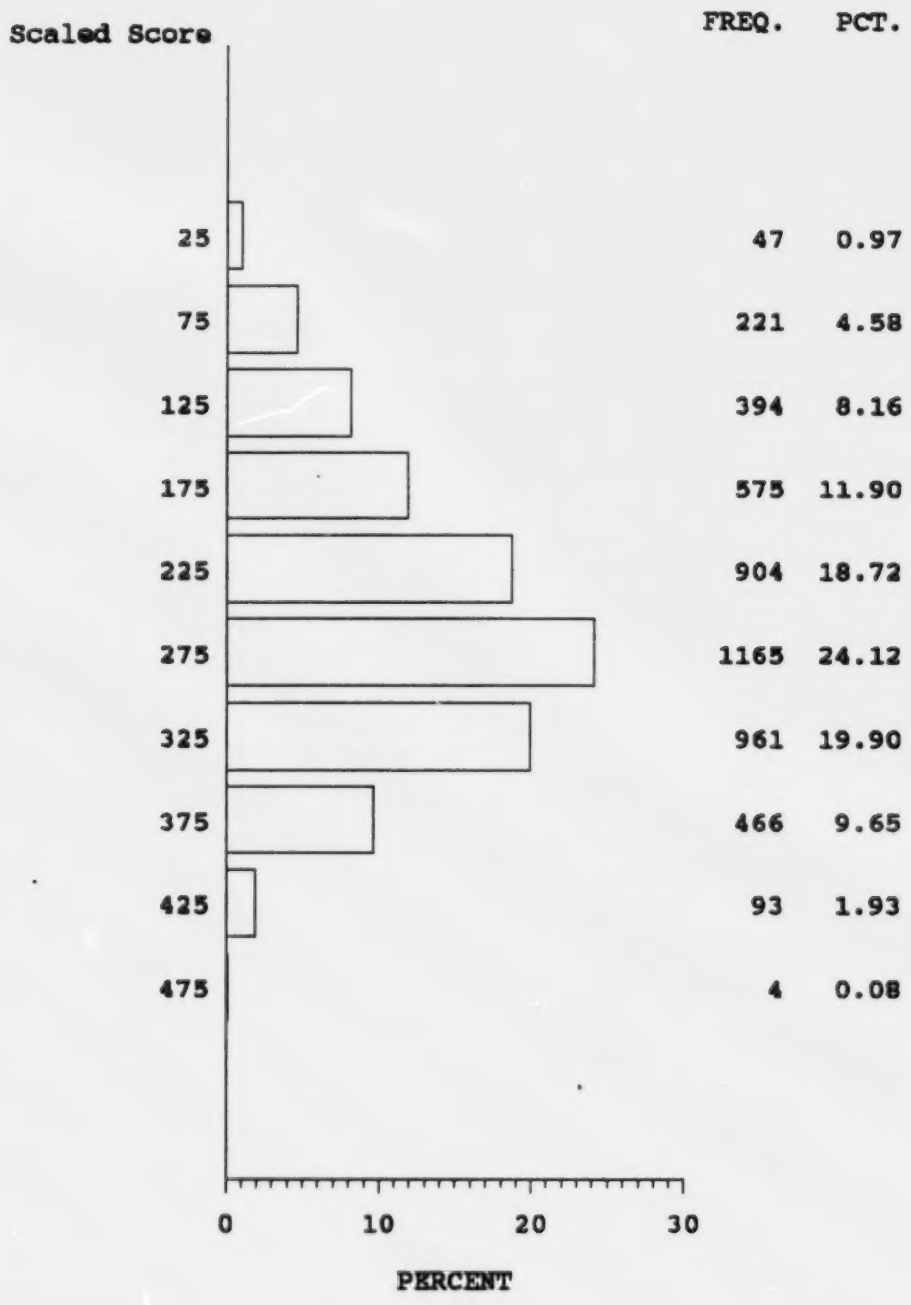

Scaled Score
FREQ.
PCT.
25
47
0.97
75
221
4.58
125
394
8.16
175
575
11.90
225
904
18.72
275
1165
24.12
325
961
19.90
375
466
9.65
425
93
1.93
475
4
0.08
0
10
20
30
PERCENT

Figure 5a

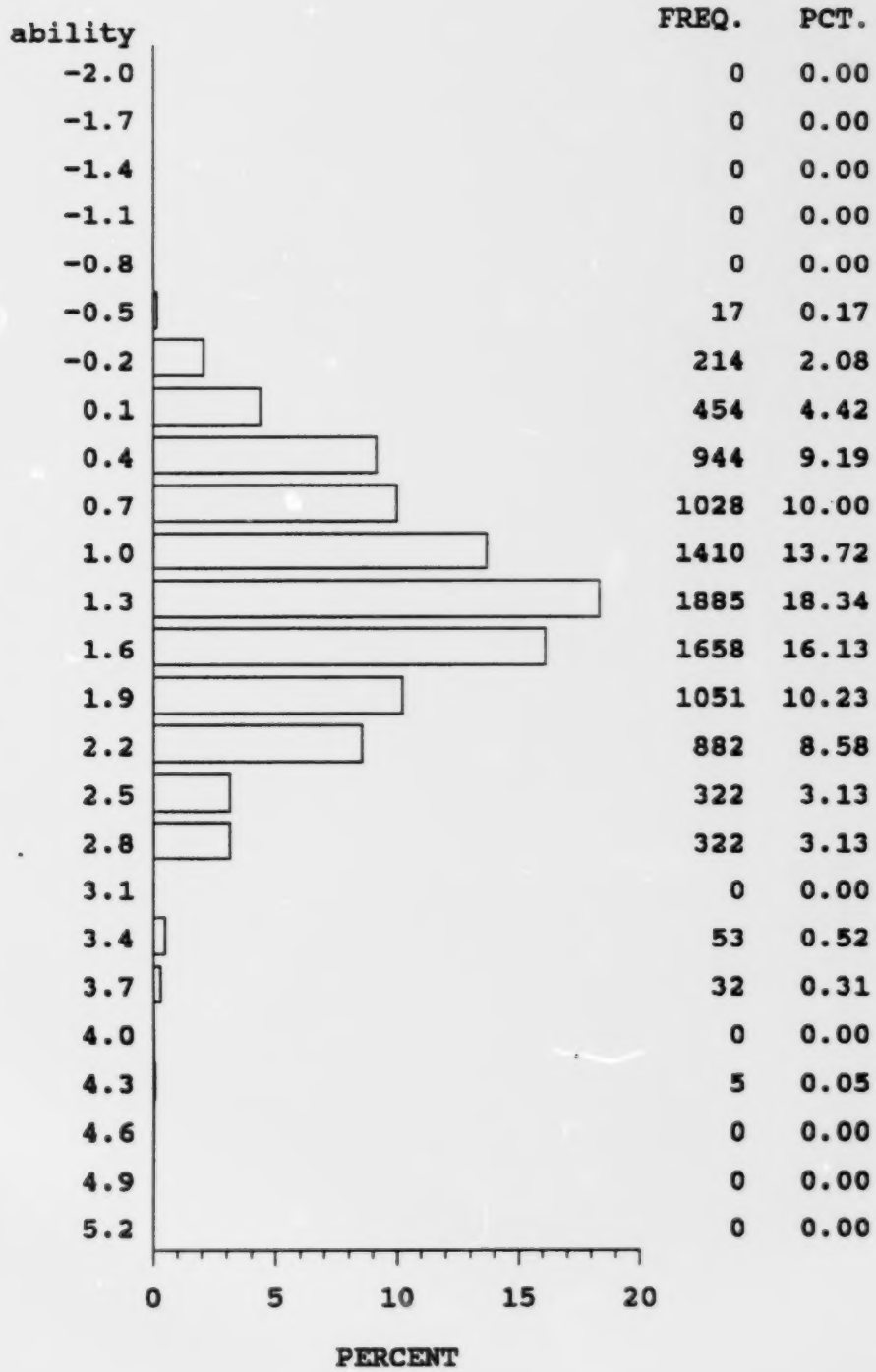

ability
FREQ.
PCT.
-2.0 0 0.00
-1.7 0 0.00
-1.4 0 0.00
-1.1 0 0.00
-0.8 0 0.00
-0.5 17 0.17
-0.2 214 2.08
0.1 454 4.42
0.4 944 9.19
0.7 1028 10.00
1.0 1410 13.72
1.3 1885 18.34
1.6 1658 16.13
1.9 1051 10.23
2.2 882 8.58
2.5 322 3.13
2.8 322 3.13
3.1 0 0.00
3.4 53 0.52
3.7 32 0.31
4.0 0 0.00
4.3 5 0.05
4.6 0 0.00
4.9 0 0.00
5.2 0 0.00
0
5
10
15
20
PERCENT

Figure 5b

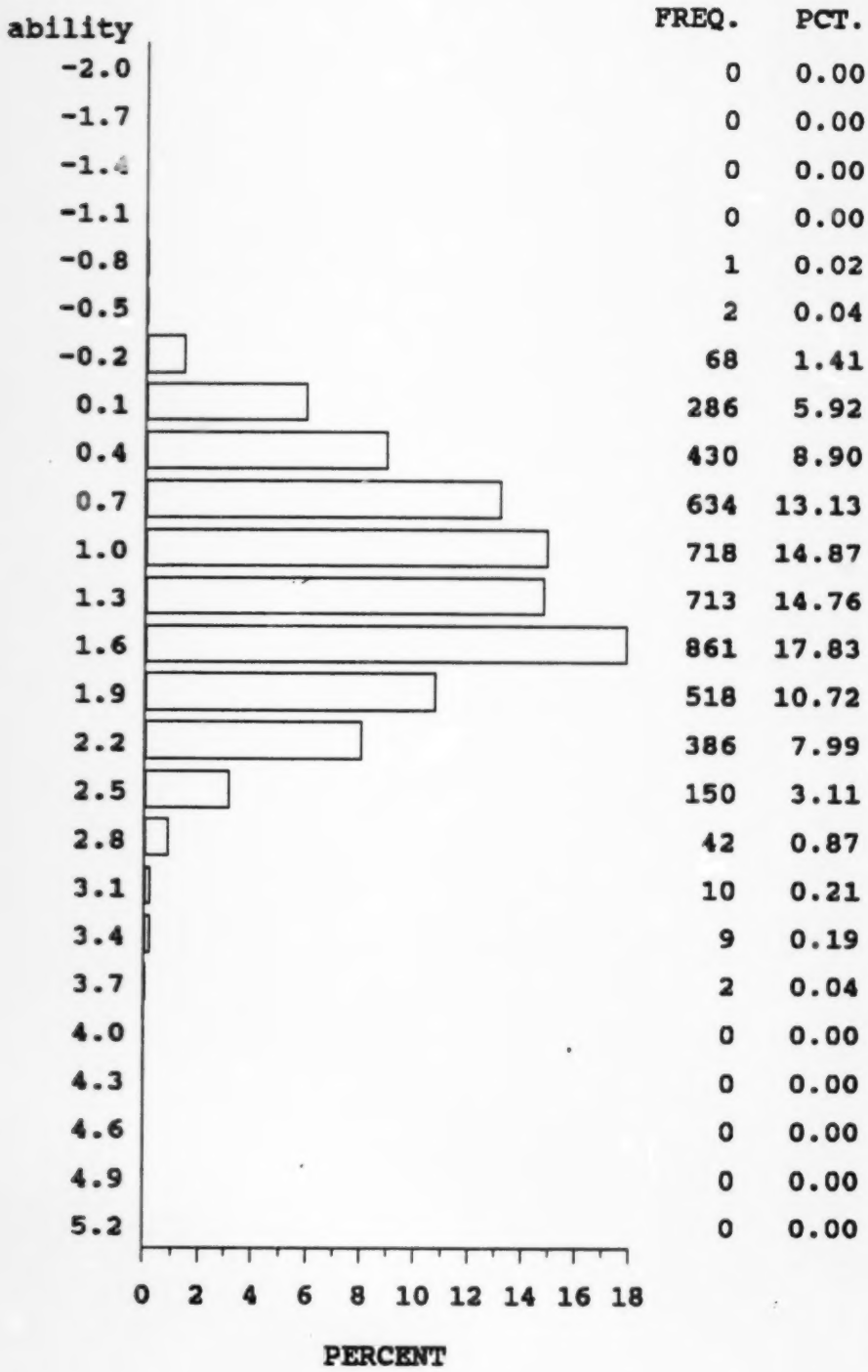

ability
FREQ.
PCT.
-2.0 0 0.00
-1.7 0 0.00
-1.4 0 0.00
-1.1 0 0.00
-0.8 1 0.02
-0.5 2 0.04
-0.2 68 1.41
0.1 286 5.92
0.4 430 8.90
0.7 634 13.13
1.0 718 14.87
1.3 713 14.76
1.6 861 17.83
1.9 518 10.72
2.2 386 7.99
2.5 150 3.11
2.8 42 0.87
3.1 10 0.21
3.4 9 0.19
3.7 2 0.04
4.0 0 0.00
4.3 0 0.00
4.6 0 0.00
4.9 0 0.00
5.2 0 0.00
0 2 4 6 8 10 12 14 16 18
PERCENT

Figure 6a

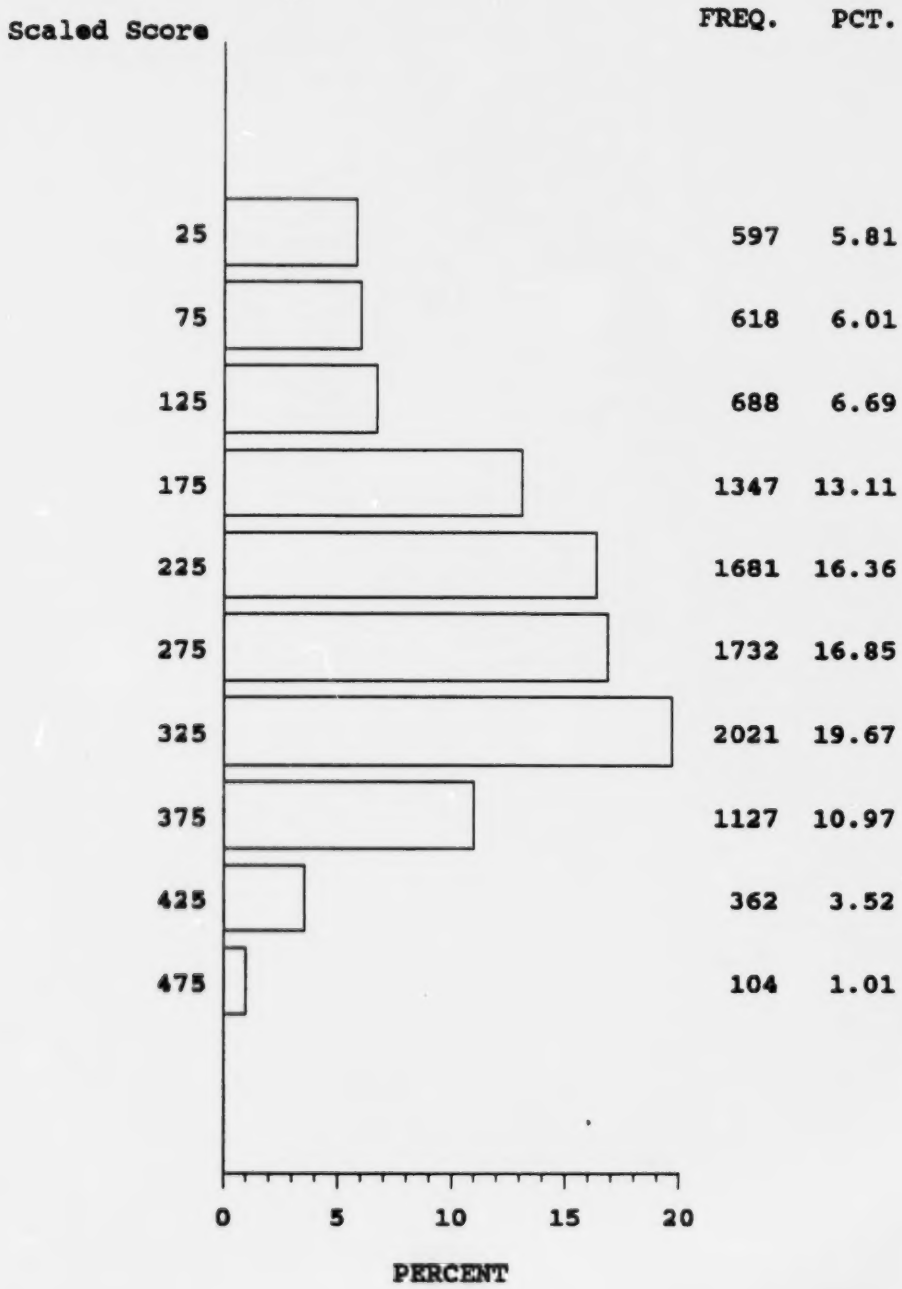

Scaled Score
FREQ.
PCT.
25
597
5.81
75
618
6.01
125
688
6.69
175
1347
13.11
225
1681
16.36
275
1732
16.85
325
2021
19.67
375
1127
10.97
425
362
3.52
475
104
1.01
0
5
10
15
20
PERCENT

Figure 6b

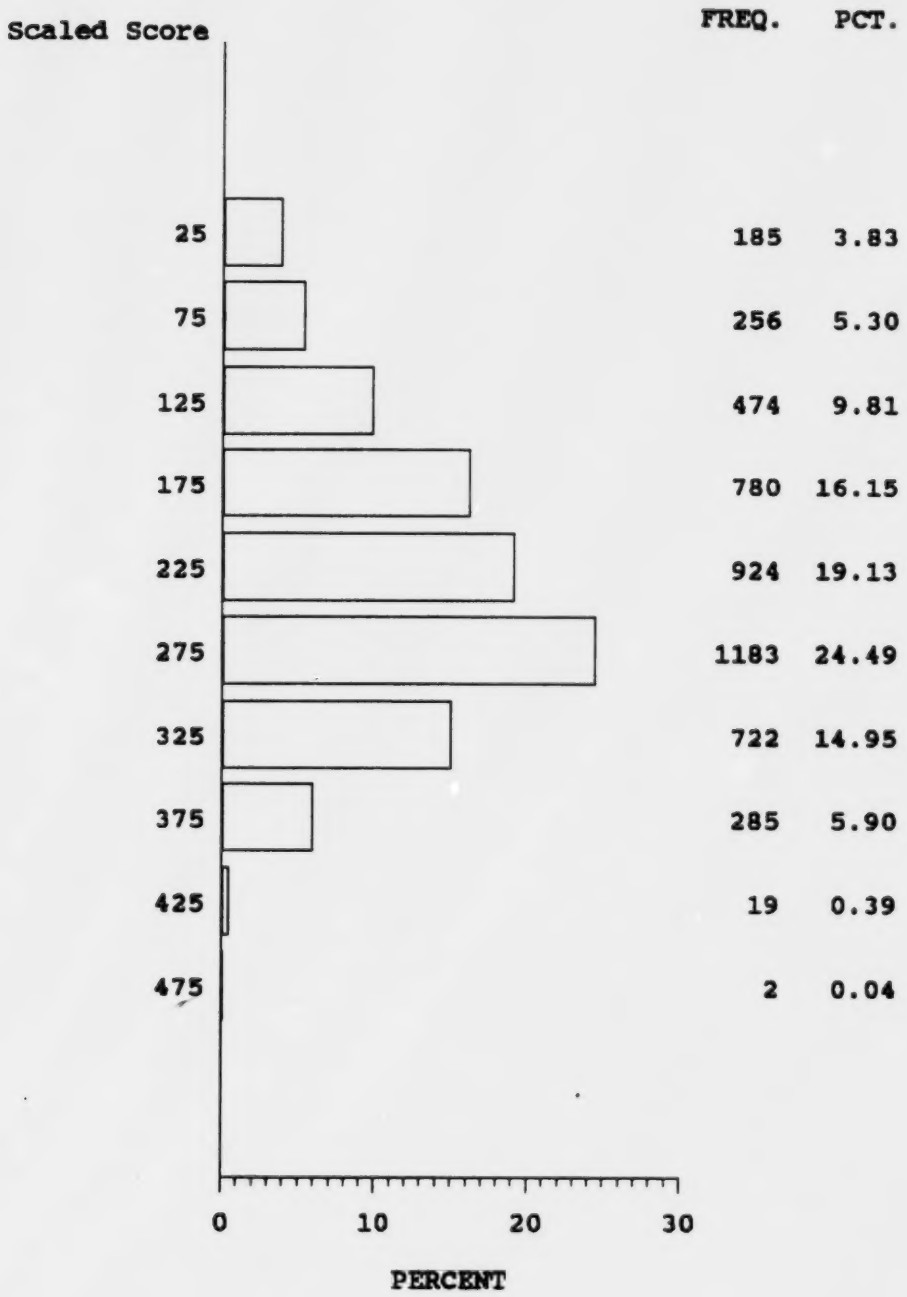

Scaled Score
FREQ.
PCT.
25
185
3.83
75
256
5.30
125
474
9.81
175
780
16.15
225
924
19.13
275
1183
24.49
325
722
14.95
375
285
5.90
425
19
0.39
475
2
0.04
0
10
20
30
PERCENT

Figure 7

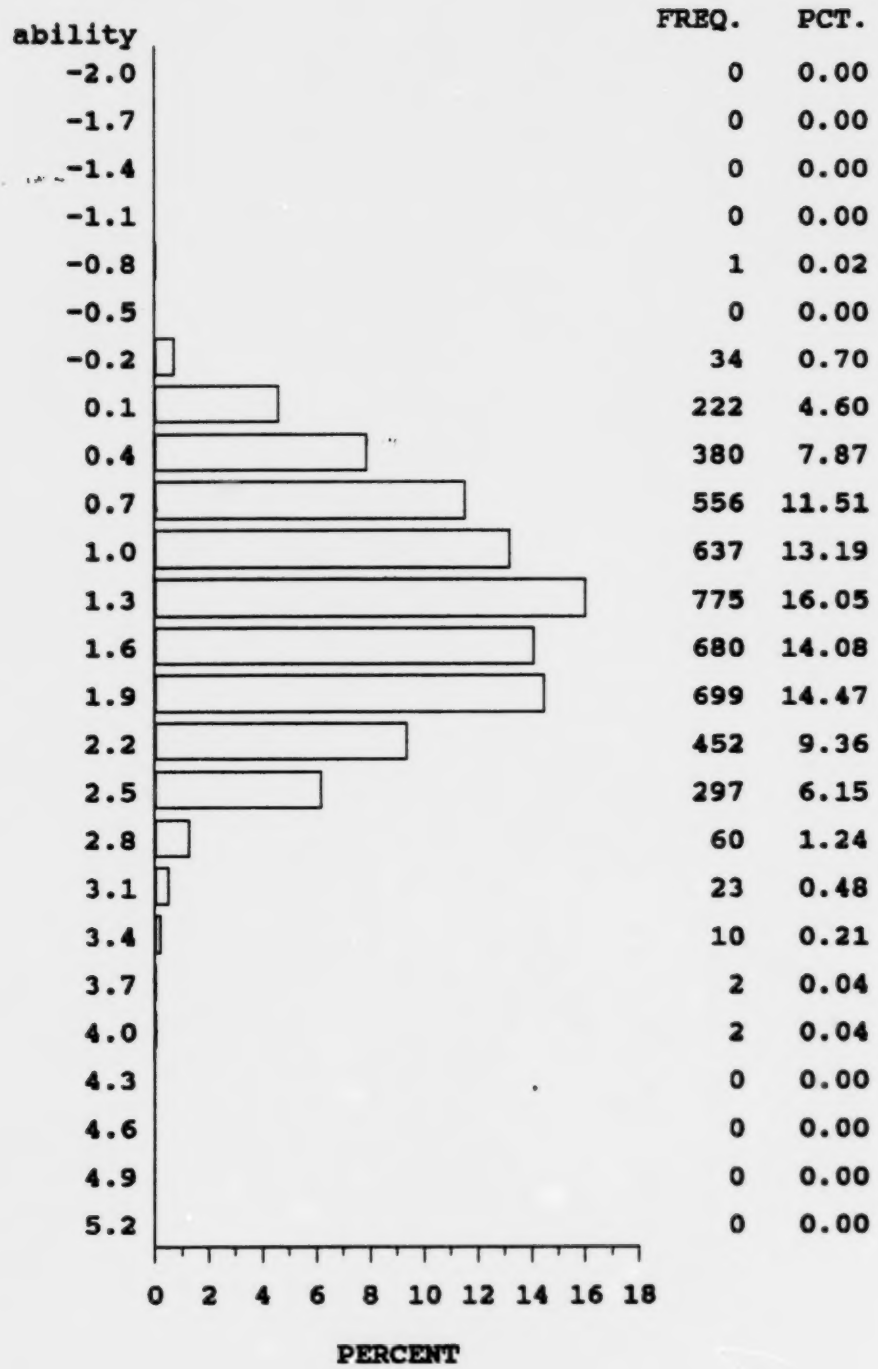

ability
FREQ.
PCT.
-2.0 0 0.00
-1.7 0 0.00
-1.4 0 0.00
-1.1 0 0.00
-0.8 1 0.02
-0.5 0 0.00
-0.2 34 0.70
0.1 222 4.60
0.4 380 7.87
0.7 556 11.51
1.0 637 13.19
1.3 775 16.05
1.6 680 14.08
1.9 699 14.47
2.2 452 9.36
2.5 297 6.15
2.8 60 1.24
3.1 23 0.48
3.4 10 0.21
3.7 2 0.04
4.0 2 0.04
4.3 0 0.00
4.6 0 0.00
4.9 0 0.00
5.2 0 0.00
0 2 4 6 8 10 12 14 16 18
PERCENT

Figure 8

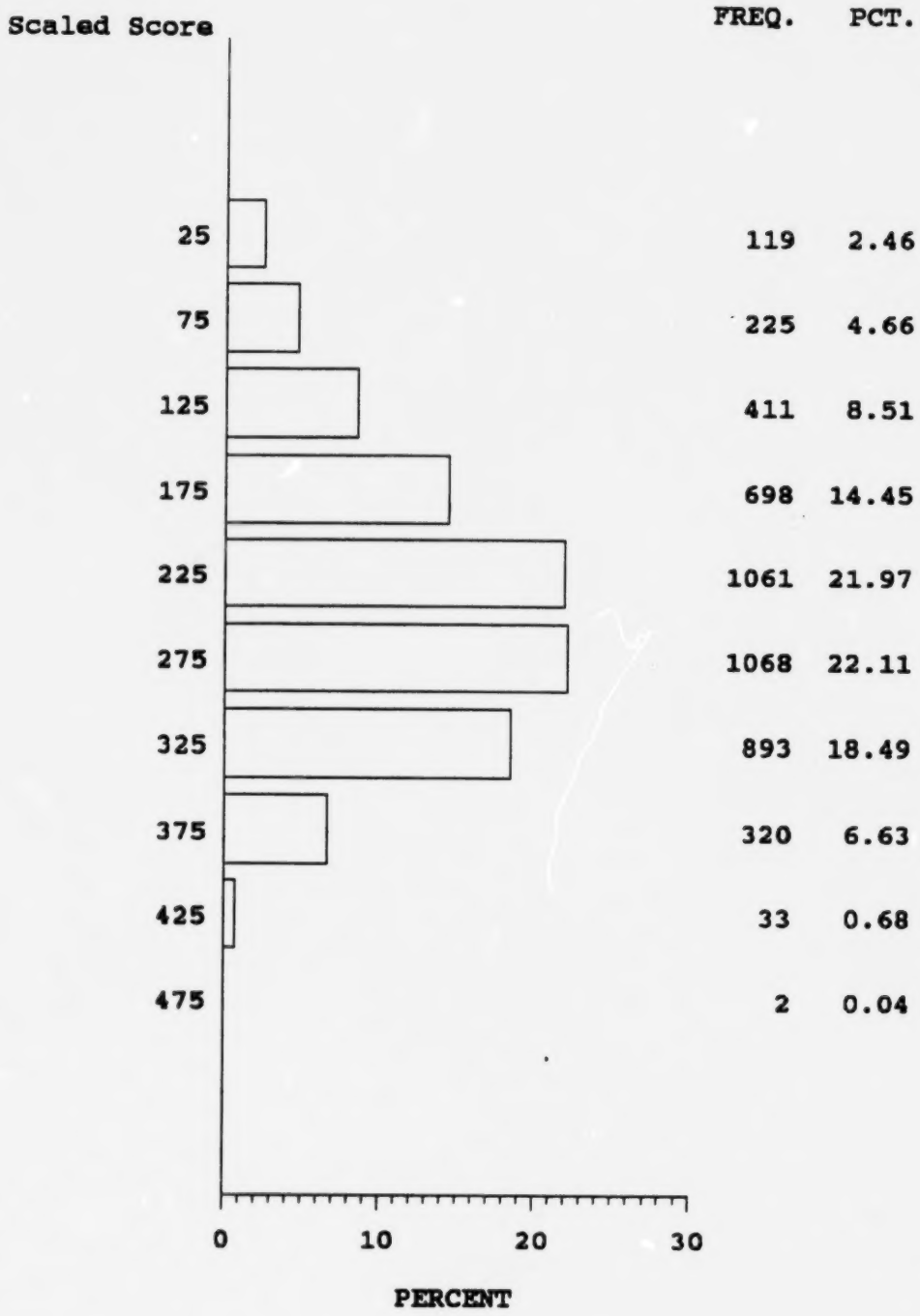

Scaled Score
FREQ.
PCT.
25
119
2.46
75
225
4.66
125
411
8.51
175
698
14.45
225
1061
21.97
275
1068
22.11
325
893
18.49
375
320
6.63
425
33
0.68
475
2
0.04
0
10
20
30
PERCENT

Figure 9

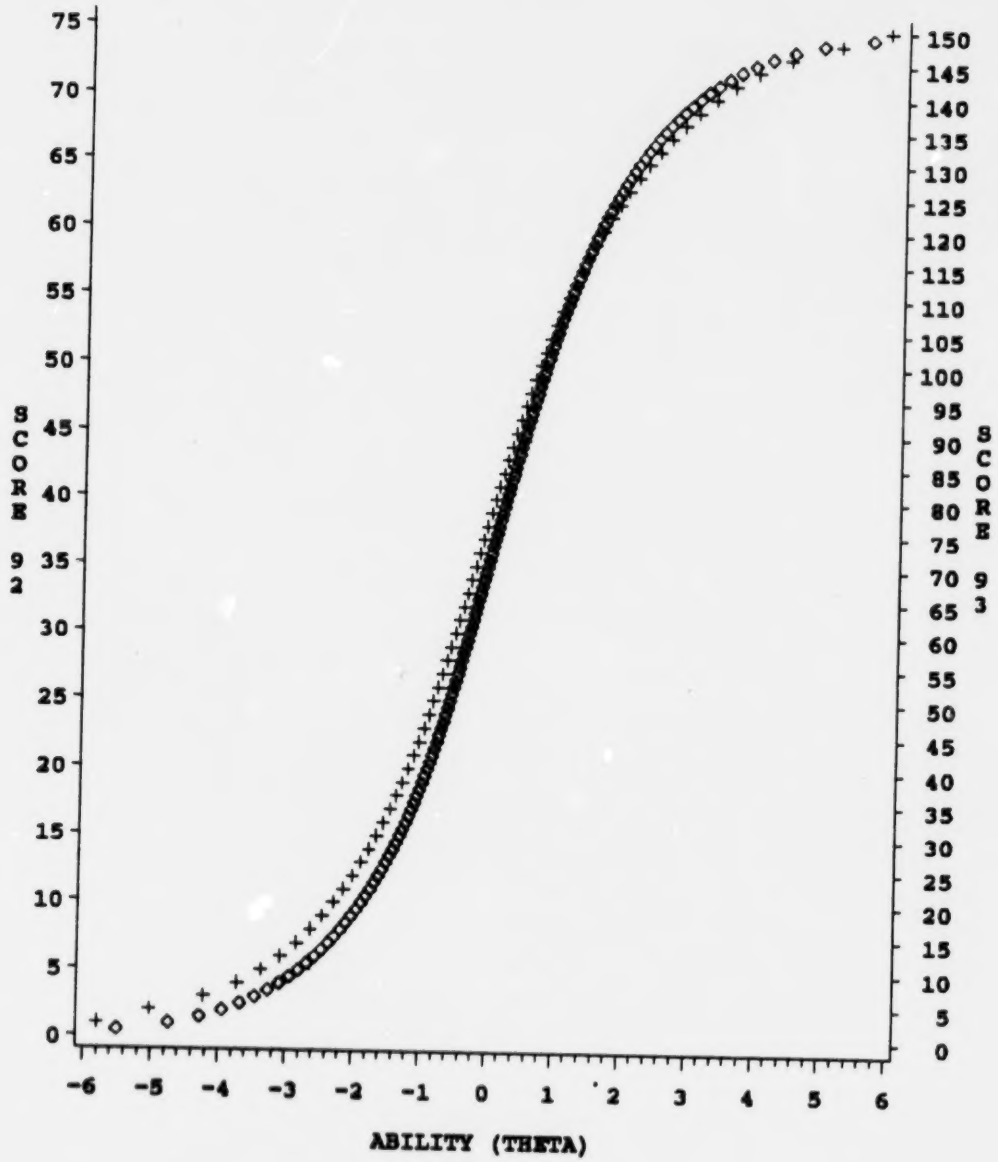


Figure 10